https://doi.org/10.51408/issi2025_121

A Novel Type Collaboration: Global Big Science Facilities Co-utilization

Zexia LI[1], Mingze Zhang[2], Lili Wang[3], Yizhan LI[4]

[1] *lizexia@mail.las.ac.cn*, [2] *zhangmingze@mail.las.ac.cn*, [4] *liyz@mail.las.ac.cn*
National Science Library, Chinese Academy of Sciences, Beijing (China)
Department of Information Resources Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing (China)

[3] *wang@merit.unu.edu*
UNU-MERIT, Maastricht University, 6211 AX Maastricht (The Netherlands)

Abstract

This paper in progress first reported a novel type of scientific collaboration, which originated from the co-utilization between or among big science facilities. 271,522 publication data was collected from 40 Synchrotron light sources (SLSs) worldwide and about 10% of the dataset is supported by more than one facility. SLSs are considered one of the most common types of big science facilities and facilitate us in reporting this novel collaboration type. Results show that from the past decades, the ratio of co-utilization has increased by about 10% but most co-utilizations are confined to two facilities. Co-utilizations might bring more scientific impact but suffer from performance loss in disruptive ability. Moreover, we discovered that most co-utilizations are user-oriented research with more authors, institutions, and knowledge input. It could also balance community participation since it could provide more chances for internal scientists, a vulnerable group in user-oriented facilities, to participate in users' research. Our progress could enrich the formality of scientific collaboration and provide a basic status of big science facility co-utilizations for reference and decision.

Introduction

Modern science is an era of big science, and the current scientific paradigm is full of collaborations, especially international research collaborations (IRC), supported by facilitated transportation and information technologies (Lin, Frey, & Wu, 2023). One of the significant features of the big science era might be knowledge convergence, caused by increasingly complex scientific issues, requiring interdisciplinary knowledge and collective wisdom (D'Ippolito & Rüling, 2019; Lauto & Valentin, 2013). Collaboration has become common for individual, institutional, and international academic entities (Katz & Martin, 1997; Wu, Wang, & Evans, 2019).

The developments of big science are highly driven by big science facilities, especially in STEM-related disciplines (Bianco, Gerhart, & Nicolson-Crotty, 2017). For the sake of giving out a better understanding of new materials, high energy physics, life science, and so on, the demands of analytical abilities in nanoscale or even more advanced are booming (Börner, Silva, & Milojevic, 2021; Heinze & Hallonsten, 2017). Such big machines are always funded by national or supranational bodies due to expensive funds, coordinative efforts, and advanced technologies (Hallonsten, 2014; Heidler & Hallonsten, 2015), but they are naturally shared with the globe to achieve the best performance in science (Söderström, 2023a). Scientists are required to submit their research proposals and await being permitted to conduct their experiments by the user commissions of the focal facility (D'Ippolito & Rüling, 2019). Therefore, scientists might travel around globally and apply for utilization chances, leading to this novel type of collaboration emerging. Collaborations between or among big science facilities are defined to originate from co-utilization in this study. Therefore, this type of scientific collaboration mainly deploys multiple experimental technologies for scientific discoveries according to the features of big science and its machines. We demonstrate that this type is novel in theory but lacks empirical evidence and would be considered a prevailing choice for scientific teams, especially in STEM-related disciplines, in modern science as demands of advanced experimental technologies increase.

This paper in progress contributes to the current literature in several ways. Firstly, the collaboration pattern could be replenished. To the authors' best knowledge, the co-utilizations of global research facilities, are initially recorded and reported. Secondly, a unique dataset, including big science facilities' publications, is collected by us, which could assist facilitymetrics to better evaluate scientific performances.

Data

There are many kinds of big science projects and research facilities, for instance, Synchrotron light sources (SLSs), Astronomical observatories, and Neutron scattering sources. SLSs are considered one of the most typical big science facilities and have been widely discussed previously. Such facilities are widely constructed around the world and broadly used in advancing knowledge in Physics, Chemistry, Medicine, Biology, and Material Sciences. Consequently, we selected SLSs in the world as cases to report this novel collaboration type.

Combined with expertise from Lightsources' staff in the Chinese Academy of Sciences and the Lightsources.org[1], we collected data from 40 global SLSs by considering the accessibility to their published records, knowledge volume, active years, and operating abilities. Their publication data are collected respectively by crawling or exporting the database on every SLS's official website from April 25 to May 10, 2024, and we only considered those publications published before 2024 and confined the document type to "article". Collecting data from the LSRIs' self-constructed databases is an accurate and credible choice (Silva, Schulz, & Noyons, 2019; Söderström, 2023a, 2023b; Söderström, Åström, & Hallonsten, 2022). The included SLSs with their locations, number of publications, and beginning year are shown in Table 1.

Notably, the numbers related to publications in Table 1 are the eventual results after the original data cleaning and matching with a bibliographical database by Python 3.11. Since most SLS databases only provide the DOI or Title of their publications, we applied the OpenAlex dataset to match more metadata for more perspective. OpenAlex is a fully open dataset, which has been widely used in previous scientometrics research (Priem, Piwowar, & Orr, 2022; Zhang et al., 2024). After data processes, the author defines the co-utilized publications as one publication that has been indexed by more than one SLS database. This criterion is also favored by

[1] https://lightsources.org/

Lightsoureces.org according to their declaration on the website and they reported about 12.5% of publications utilized more than one facility[2].

From Table 1, the involved SLSs mainly located in developed nations or regions. Some developing nations or regions also constructed SLSs, Armenia, Brazil, China, and Jordan but their participation ratios of co-utilization are not as well as their developed counterparts.

Table 1. Published Records Distribution Among All Synchrotron Light Sources.

No.	SLS	C/R	BY	NP	NCP	NCP/NP (%)
1	Center for the Advancement of Natural Discoveries using Light Emission (CANDL)	Armenia	2013	121	5	4.132
2	Australian Synchrotron (AS)	Australia	2006	7,000	1,048	14.971
3	Laboratório Nacional de Luz Síncrotron (LNLS)	Brazil	1985	4,903	306	6.241
4	Canadian Light Source (CLS)	Canada	2006	4,339	1,347	31.044
5	Beijing Synchrotron Radiation Facility (BSRF)	China	1992	5,106	1,492	29.221
6	National Synchrotron Radiation Laboratory (NSRL)	China	1971	6,513	1,258	19.315
7	Shanghai Synchrotron Radiation Facility (SSRF)	China	2000	10,451	2,153	20.601
8	Institute for Storage Ring Facilities (ISRF)	Denmark	1983	982	163	16.599
9	European Synchrotron Radiation Facility (ESRF)	France	1979	33,351	5,894	17.673
10	SOLEIL	France	2005	5,758	1,624	28.204
11	KIT Light Source (KIT)	Germany	2014	674	226	33.531
12	BESSY II - Helmholtz-Zentrum Berlin (BESSY)	Germany	1992	7,347	1,640	22.322
13	Dortmund Electron Storage Ring Facility (DESRF)	Germany	2009	312	61	19.551
14	Electron Stretcher Accelerator (ELSA)	Germany	1968	83	1	1.205
15	Metrology Light Source (MLS)	Germany	1957	8,943	379	4.238
16	PETRA III at DESY (PETRA)	Germany	1950	31,672	3,634	11.474
17	DAFNE	Italy	2010	45	5	11.111
18	Elettra Synchrotron Light Laboratory (ELETTRA)	Italy	1994	6,521	1,082	16.593
19	Aichi Synchrotron Radiation Center (ASRC)	Japan	2014	58	9	15.517
20	Hiroshima Synchrotron Radiation Center (HSRC)	Japan	2008	329	95	28.875
21	Photon Factory (PF)	Japan	1969	14,518	2,239	15.422
22	Ritsumeikan University SR Center (RUSRC)	Japan	2009	218	55	25.229
23	Saga Light Source (SAGA)	Japan	2004	257	39	15.175
24	SPring-8	Japan	1999	16,209	2,719	16.775
25	Ultraviolet Synchrotron Orbital Radiation Facility (USORF)	Japan	1997	737	102	13.840
26	Synchrotron-light for Experimental Science and Applications in the Middle East (SESAME)	Jordan	2012	86	18	20.930

[2] https://lightsources.org/about-2/

27	Pohang Light Source (PLS)	Korea	2008	6,012	339	5.639
28	National Synchrotron Radiation Centre (SOLARIS)	Poland	2018	210	38	18.095
29	Kurchatov Synchrotron Radiation Source (KSRS)	Russia	2004	282	32	11.348
30	Singapore Synchrotron Light Source (SSLS)	Singapore	2015	174	24	13.793
31	ALBA	Spain	2005	2,470	749	30.324
32	MAX IV Laboratory (MAXIV)	Sweden	1982	4,655	874	18.776
33	Swiss Light Source (SLS)	Switzerland	2007	1,438	358	24.896
34	National Synchrotron Radiation Research Center (NSRRC)	Taiwan (China)	2003	6,783	986	14.536
35	Diamond Light Source (DIAMOND)	United Kingdom	1983	13,114	3,125	23.829
36	Advanced Light Source (ALS)	USA	1991	16,764	3,709	22.125
37	Advanced Photon Source (APS)	USA	1970	31,326	5,464	17.442
38	Cornell High Energy Synchrotron Source (CHESS)	USA	1997	1,228	290	23.616
39	National Synchrotron Light Source II (NSLSII)	USA	1984	12,302	2,504	20.354
40	Stanford Synchrotron Radiation Lightsource (SSRL)	USA	1983	8,231	2,498	30.349
	Total Data			245,984	23,046	9.37

Note: C/R: Country/Region; BY: Begin Year; NP: Number of Publications; NCP: Number of Co-utilized Publications; Alphabet Order by the Column: LC/R; NP-Total Data and NCP-Total Data has been de-duplicated by WorkID in OpenAlex.

Progress

Current Status of Co-utilizations

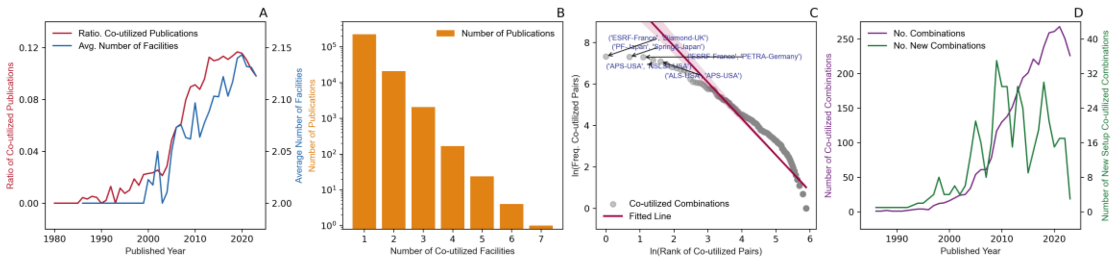

Figure 1. Current Status of Co-utilization.

Figure 1(A) displays the annual ratio distribution of co-utilized published records in red color and the average number of co-utilized facilities in blue line. The ratio of co-utilized publications increased from zero to ten percentile as time went on, and gradually more global facilities participated in co-utilization since the average number of facilities is observed increasing. A similar trend could also be observed in Figure 1(D) that the annual combinations of big science facilities are also increasing (purple color), and, each year, new combinations are set up (green color). However, these booming trends declined after 2020, which might be influenced by the time lag of self-constructed databases and the COVID-19 pandemic, especially the following quarantine time and travel restrictions. In total, co-utilization has shown increasing trends in the past and might keep booming in the future.

The number distribution of publications related to the number of co-utilized facilities is shown in Figure 1(B). The number of co-utilized facilities increases by one unit, the number of publications might receive a tenfold decline approximately. In Figure 1(C), we recorded those highly frequent combinations and applied a linear fit to the distribution, contributing to describing the mechanism of facilities co-utilization. In the figure, almost every top choice shows great preference in geography that the facilities in the combinations might be in the same nation or region, for instance: both *PF* and *SPring-8* are Japanese facilities; *APS*, *NSLS-II*, and *ALS* are in the USA; *ESRF*, *Diamond*, and *PETRA* are in Europe. In total, more combinations involved might be a future trend and it is important to unveil the relationship between novel or common combinations and scientific breakthroughs and understand the impact of global technological co-utilization. In Figure 2, we could also observe the impacts of geographical factors in North America, Europe, and East Asia.

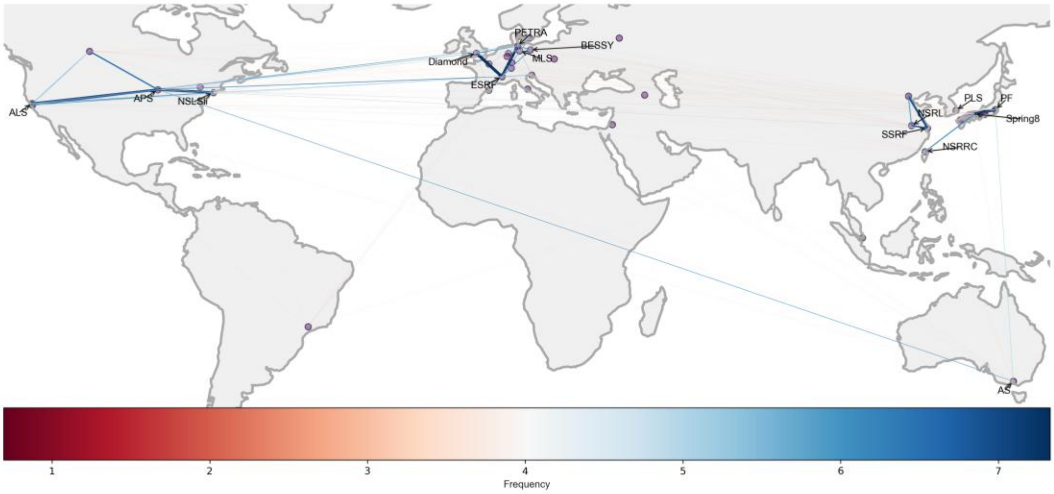

Figure 2. Global Distributions of Co-utilized Facilities.

We visualized the co-utilized relationships between global big science facilities and enclosed the names of the Top 15 facilities in productivity for better indication in Figure 2. The nodes in the figure represent big science facilities in our dataset and the links represent the frequency of co-utilizations between every two facilities with observations.

Potentially differences between Co-utilization and Singly utilization.

We adopted the Disruption Index (DI) as an indicator to measure the disruptive performance of scientific publications. DI was proposed by Funk and Owen-Smith (2017) and revised by Wu et al. (2019), and it has been widely used in scientometrics. Limited by the pages, we do not introduce this indicator in this progress work. In brief, if DI>0, indicating that the focal paper might bring a new orientation in knowledge system while DI<0, the focal paper might consolidate the current

knowledge system (Lin et al., 2023; Zhang et al., 2024). We mainly used the probability of disruption and considered disruptive publications are DI>=0.
Additionally, in the context of user-oriented big science facilities, there are two main research communities, External Scientists (Users, who visit the facility) and Internal Scientists (Staff, affiliated with the facilities), and the users are always in domination and the staff might be overlooked in the scientific publications since users might collaborate but will not co-author with them (D'Ippolito & Rüling, 2019; Söderström, 2023b). However, we demonstrate that co-utilization might bring more chances for internal scientists to co-author in user research.

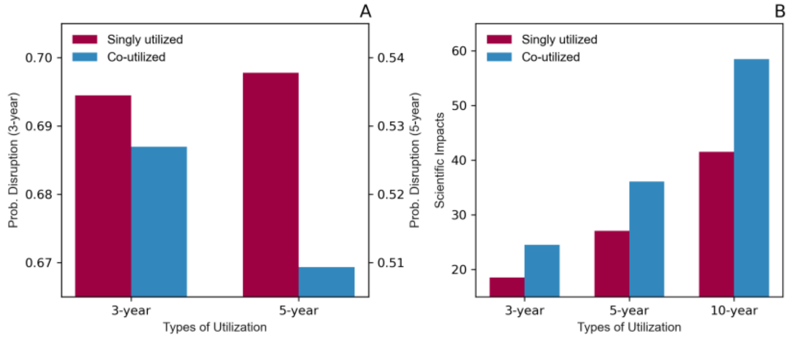

Figure 3. Performance Differences in Disruption and Scientific Impacts between Co-utilization and Singly Utilization.

In Figure 3, we mainly displayed the performance gaps between co-utilizations and single utilization by measuring the disruptive probability (A) and scientific impacts (B) of their supporting publications. Singly utilizations might produce more disruptive knowledge but receive fewer citations than co-utilizations since published in a 3-year, a 5-year, and a 10-year citation window.

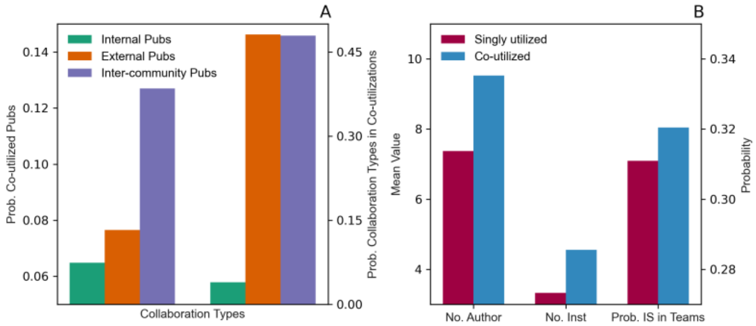

Figure 4. Differences between Co-utilizations and Single utilizations.

In Figure 4(A), we observe that above 12% of inter-community publications are supported by more than one facility (co-utilization) and the value is much higher than the ratio (above 7%) in External publications (authored by external scientists at all). Moreover, in the dataset of co-utilizations (23,046 papers are mentioned in Table 1), the ratios of inter-community publications and external publications are close, which also reveals that co-utilizations might provide more chances for staff participation. In Figure 4(B), we demonstrate that co-utilization might involve more authors and institutions in collaboration and the probability of internal scientists participating in teams is also higher than single utilization, which further ensures that co-utilization might balance the community participation.

Conclusion and Future Works

This research in progress mainly reports a novel type of scientific collaboration based on a unique dataset of publications collected by us by crawling or exporting bibliography from SLSs' self-constructed databases. Future works could further explore the relationships between co-utilizations and scientific performance in the context of a resource-based view and the theory of S&T human capital. Moreover, we would also compare the main differences between facility co-utilization and inter-organizational collaboration in academia.

Acknowledgments

This work was supported by the National Social Science Fund Major Projects of China (Project No. 22&ZD127). We would like to thank Lingling Zhang, Yuhui Dong, and Honghong Li for their expertise and assistance with the basic knowledge of Large-scale Research Infrastructures and valuable comments from reviewers.

References

Bianco, W., Gerhart, D., & Nicolson-Crotty, S. (2017). Waypoints for Evaluating Big Science*. *Social Science Quarterly, 98*(4), 1144-1150. doi:10.1111/ssqu.12467

Börner, K., Silva, F. N., & Milojevic, S. (2021). Visualizing big science projects. *Nature Reviews Physics, 3*(11), 753-761. doi:10.1038/s42254-021-00374-7

D'Ippolito, B., & Rüling, C. C. (2019). Research collaboration in Large Scale Research Infrastructures: Collaboration types and policy implications. *Research Policy, 48*(5), 1282-1296. doi:10.1016/j.respol.2019.01.011

Funk, R. J., & Owen-Smith, J. (2017). A Dynamic Network Measure of Technological Change. *Management Science, 63*(3), 791-817. doi:10.1287/mnsc.2015.2366

Hallonsten, O. (2014). How expensive is Big Science? Consequences of using simple publication counts in performance assessment of large scientific facilities. *Scientometrics, 100*(2), 483-496. doi:10.1007/s11192-014-1249-z

Heidler, R., & Hallonsten, O. (2015). Qualifying the performance evaluation of Big Science beyond productivity, impact and costs. *Scientometrics, 104*(1), 295-312. doi:10.1007/s11192-015-1577-7

Heinze, T., & Hallonsten, O. (2017). The reinvention of the SLAC National Accelerator Laboratory, 1992-2012. *History and Technology, 33*(3), 300-332. doi:10.1080/07341512.2018.1449711

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26*(1), 1-18. doi:10.1016/s0048-7333(96)00917-1

Lauto, G., & Valentin, F. (2013). How Large-Scale Research Facilities Connect to Global Research. *Review of Policy Research, 30*(4), 381-408. doi:10.1111/ropr.12027

Lin, Y., Frey, C. B., & Wu, L. (2023). Remote collaboration fuses fewer breakthrough ideas. *Nature, 623*(7989), 987-991. doi:10.1038/s41586-023-06767-1

Priem, J., Piwowar, H. A., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *ArXiv, abs/2205.01833*.

Silva, F. S. V., Schulz, P. A., & Noyons, E. C. M. (2019). Co-authorship networks and research impact in large research facilities: benchmarking internal reports and bibliometric databases. *Scientometrics, 118*(1), 93-108. doi:10.1007/s11192-018-2967-4

Söderström, K. R. (2023a). Global reach, regional strength: Spatial patterns of a big science facility. *Journal of the Association for Information Science and Technology, 74*(9), 1140-1156. doi:10.1002/asi.24811

Söderström, K. R. (2023b). The structure and dynamics of instrument collaboration networks. *Scientometrics, 128*(6), 3581-3600. doi:10.1007/s11192-023-04658-w

Söderström, K. R., Åström, F., & Hallonsten, O. (2022). Generic instruments in a synchrotron radiation facility. *Quantitative Science Studies, 3*(2), 420-442. doi:10.1162/qss_a_00190

Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature, 566*(7744), 378-382. doi:10.1038/s41586-019-0941-9

Zhang, M.-Z., Wang, T.-R., Lyu, P.-H., Chen, Q.-M., Li, Z.-X., & Ngai, E. W. T. (2024). Impact of gender composition of academic teams on disruptive output. *Journal of Informetrics, 18*(2), 101520. doi:https://doi.org/10.1016/j.joi.2024.101520